SINET: Ethiop. J. Sci., 28(1):1–14, 2005
 ISSN: 0379–2897

INVERSE GAUSSIAN MODEL FOR SMALL AREA ESTIMATION VIA GIBBS SAMPLING

Fassil Nebebe [1], Cynthia M. DeSouza [2] and Yogendra P. Chaubey [3]

[1] Department of Decision Sciences and MIS, Concordia University, Montréal, Québec H3G 1M8 Canada. E-mail: fnebebe@vax2.concordia.ca
[2] Medtronic, Inc., MS T278, Tachy Clinical Division, Minneapolis, MN 55432 USA
[3] Department of Mathematics and Statistics, Concordia University, Montréal, Québec H4B 1R6 Canada

ABSTRACT: We present a Bayesian method for estimating small area parameters under an inverse Gaussian model. The method is extended to estimate small area parameters for finite populations. The Gibbs sampler is proposed as a mechanism for implementing the Bayesian paradigm. We illustrate the method by application to household income survey data, comparing it against the usual lognormal model for positively skewed data.

Key words/phrases: Finite population sampling, hierarchical Bayesian inference, lognormal model, MCMC integration, shrinkage estimates

INTRODUCTION

Small domain estimates are required by policy makers for a diversity of subpopulations in order to make decisions on issues relating to small areas. These small domains need not be geographical locations, but can represent distinct subdomains defined by several stratification factors. Sample survey data are available for a number of small domains, cross-classified by non-overlapping and exhaustive subgroups of the population, requiring estimates for small areas and the corresponding interest in methods for producing such estimates. Various branches of governments have been involved in research to obtain small area estimates for use in decision making in local areas. Examples of some of the research include, studies of per capita income for states and local government areas (Fay and Herriot, 1979); estimates of crop yields, population counts and unemployment rates (Schaible, 1996); and studies of health needs (Malec *et al.*, 1999).

Several authors have also considered the problem of small area estimation from various perspectives primarily using the Gaussian model and the classical techniques of estimation.

An early work by Purcell and Kish (1979) presented a comprehensive review of sample survey research in small area estimation. Ghosh and Meeden (1986) introduced an empirical Bayes approach in normal model finite population sampling theory for small area estimation. Ghosh and Rao (1994) and Rao (1999) presented accessible reviews of several of the techniques for small area estimation and indicated the advantages of the Bayesian and empirical Bayes approaches over the classical methods. In a recent book, Rao (2003) provides details of various methods of estimation, the wide range of available models and issues associated with small area studies.

The importance of non-normal models in small area estimation has also been investigated by some authors. For example, MacGibbon and Tomberlin (1989) have considered estimating small area rates and binomial parameters using empirical Bayes methods. Stroud (1991) used hierarchical Bayes approach for univariate natural exponential families with quadratic variance functions in sample survey applications, while Chaubey *et al.* (1994) extended the work by Fries and Bhattacharyya (1983) to include the maximum likelihood analysis of the two-factor inverse Gaussian model for the unbalanced and interaction case for the estimation of small area parameters in finite populations.

The object of this article is to develop a Bayesian approach for small area estimation under an inverse Gaussian model, denoted Inverse Gaussian (q, s^2), whose density function is given by

$$f(y;\ q,\ s^2) = (2ps^2)^{-1/2} y^{-3/2}\ exp\{-(yq^{-1}-1)^2/2s^2\},$$
$$y > 0,\ q > 0,\ s^2 > 0,$$

with $E(y) = q$, and $V(y) = ?^3 s^2$. Note that neither $?$ nor s^2 is a location or scale parameter under this model. A reciprocal-linear model for the factor effects is motivated from the context of the

underlying Wiener process (see Bhattacharyya and Fries, 1982). Specifically, if $Y(t)$ is a Wiener process with drift $\mu > 0$, then the random time when the process first hits a specified positive barrier, has an inverse Gaussian distribution (Cox and Miller, 1965). The interpretation of the inverse Gaussian distribution as a first passage time distribution suggests its potential usefulness in modeling lifetimes as an alternative to other conventional models such as the Weibull, gamma and lognormal.

We consider the common two-factor model $y_{ijk} = \mu_{ij} + e_{ijk}$, $k = 1, \ldots, n_{ij}$, where the e_{ijk} are independent error terms having an inverse Gaussian density with mean zero, and the μ_{ij} are unknown mean parameters. A Bayesian analysis involves a prior specification for the μ_{ij} conditional on hyperparameters ?, i.e., $\mu_{ij} = g(?) + t_{ij}$, where $g(?)$ is a prior mean parameter and the t_{ij} are random errors distributed independently of the e_{ijk} according to a density p with mean zero. Adaptive Markov Chain Monte Carlo (MCMC) integration methods, such as the Gibbs sampler (Gelfand and Smith, 1990), have proved to be a powerful tool for analyzing conjugate Bayesian hierarchical models. The Bayesian paradigm allows us to use information from similar sources in constructing estimators and predictors, in addition to the most directly available source of information. This is extremely useful in small area estimation theory, where one can improve the estimates in a certain area by incorporating information from similar neighboring areas.

In the next few sections, we will discuss our Bayesian model specification and provide a general paradigm for the Bayesian modeling of positively skewed data using an inverse Gaussian model. This model is compared with the usual lognormal model. We illustrate our approach by application to a household income data obtained from Statistics Canada (1987). The data set contains comprehensive information on each household, such as number of persons, number of adults, type of dwelling, educational level of the head of household, etc. We choose the domains of the study as the ten regions stratified by six education classes. The principal characteristic of interest is household income. Although the problem is characterized as a small area estimation, the discussion can apply to any stratified random sampling design considered for estimation at the stratum level.

THE MODEL

Consider a population U divided into J nonoverlapping small areas labelled $j = 1, \ldots, J$, and a second classification of the population into I groups labelled $i = 1, \ldots, I$. The total sample n is thereby cross-classified into IJ nonoverlapping cells of size n_{ij}, $n = S_{ij}\, n_{ij}$. In practice a simple random sample is drawn from the entire population, so that the n units are post classified into the cells, and the cell counts n_{ij} are random. We will assume that a stratified random sample design is used such that each cell defines a stratum from which a random sample of size n_{ij} is drawn. Following the terminology of a two-factor model in factorial experiments we let I denote the number of levels of the row factor A and J denote the number of levels of the column factor B. At each factor setting or cell (i,j), a sample of size n_{ij} is selected.

We focus on the additive or no-interaction model which assumes that the drift of the Wiener process corresponding to each cell is the sum of the factor effects. Since the mean is inversely proportional to the drift of the Wiener process, the usual parameterization of additive effects suggests the following model:

$y_{ijk} \sim$ Inverse Gaussian(q_{ij}, s^2), $k = 1, \ldots, n_{ij}$,

$$q_{ij}^{-1} = a_i + b_j,\ q_{ij} > 0, s^2 > 0, \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

with $E(y_{ijk}) = \theta_{ij}$, $V(y_{ij}) = q_{ij}^3\ s^2$. Let y denote the collection of data over all IJ domains. The main problem of interest is to combine the data y with prior information about the unknown parameters $f = (\{a_i\}, \{b_j\}, s^2)$ and obtain their posterior distributions. By Bayes theorem, for a given prior pdf $\pi(\phi)$, the posterior of interest is given by $\pi(f/y) \propto \pi(\phi) f(y/f)$. Other questions of interest include making inferences about functionals of the parameters, such as the predictive density of future y_{ijk}'s. Due to the nonlinearity arising from the parametrization of the model, these posterior computations are intractable if approached directly via Bayes theorem. However, the model has a

convenient conditional structure that lends itself to the method of Gibbs sampling.

Gibbs sampling

A series of papers discuss Gibbs sampling for conjugate Bayesian models and the calculation of marginal posterior densities and moments (see Gelfand and Smith, 1990; 1991; Gelfand *et al*., 1990; 1992; Gelfand and Smith, Casella and George, 1992). It is assumed that our collection of random variables are such that specification of all full conditional densities uniquely determines the full joint density. Then the Gibbs sampler is an iterative Monte Carlo method designed to extract marginal densities from these full conditional densities. Consider three variables θ_1, q_2, q_3 with joint density $[\theta_1, q_2, q_3]$. Suppose the full conditional densities are $[\theta_1, |q_2, q_3]$, $[\theta_2, |q_1, q_3]$, and $[\theta_3, |q_1, q_2]$. After M cycles of iterations, the sample $(q_1^{(M)}, q_2^{(M)}, q_3^{(M)})$ is obtained. Under suitable regularity conditions, as $M \rightarrow 8$, the sampled values converge in distribution to the relevant marginal and joint distributions, *i.e*., $q_j^{(M)} \rightarrow q_j \sim [q_j]$, and $(q_1^{(M)}, q_2^{(M)}, q_3^{(M)}) \rightarrow (q_1, q_2, q_3) \sim [q_1, q_2, q_3]$. For M large, so that the desired convergence in distribution has been attained, the N values $(q_1^{(j)}, q_2^{(j)}, q_3^{(j)})$, j=$M$+1,..., M+N are Markov sample from $[\theta_1, q_2, q_3]$. In order to cover a significant portion of the space generated by the posterior density, Gelman and Rubin (1992) recommend the use of multiple runs wherein the sampler is replicated with different starting values and the M^{th} iterate from each run is retained.

The burn-in length M, is dependent on starting values and the convergence rates of algorithm to a stationary distribution depends on the targeted posterior. Several approaches to handle these problems have been suggested in the literature (see e.g., Cowles and Carlin, 1996; Roberts and Rosenthal, 1998, and the references cited there). However, in single as well as multiple runs, posterior inference is straightforward since the entire posterior sample is available. For instance, the marginal density of θ_1 is obtained as a finite mixture,

$$[\hat{q}_1] = N^{-1} \sum_{t=M+1}^{M+N} [q_1 | q_2^{(t)}, q_3^{(t)}] \quad \text{.....(2)}$$

The expectation of a function $g(\theta)$ of the parameters is estimated via the sample average

$$\hat{E}(g(q)) = N^{-1} \sum_{t=M+1}^{M+N} g(q^{(t)}). \quad \text{.....(3)}$$

It is possible to improve our estimation by taking advantage of the full conditional densities. If for any s, the conditional expectation $g_s(q_r, r \neq s) = E(g(q) | q_r, r \neq s)$ is available in closed form, then by a Rao-Blackwell argument, an estimator with smaller mean squared error is

$$\hat{g}_s = N^{-1} \sum_{t=M+1}^{M+N} g_s(q_r^{(t)}, r \neq s). \quad \text{.....(4)}$$

With regard to density estimation, again if for any s, θ_s appears as an argument of f, the conditional density $[f|\theta_r, r \neq s]$ can be obtained by a univariate transformation from $[\theta_s|\theta_r, r \neq s]$. The resulting Rao-Blackwellized sample-based density estimate of f is

$$[\hat{f}]_s = N^{-1} \sum_{t=1}^{N} [f | q_r^{(t)}, r \neq s]. \quad \text{.....(5)}$$

Full conditional distributions

The likelihood function for ($\{\alpha_i\}$, $\{\beta_j\}$, σ^2) under the inverse Gaussian model is

$$(2ps^2)^{-\frac{n}{2}} \Pi_i \Pi_j \Pi_k y_{ijk}^{-\frac{3}{2}} \exp(-(\Sigma_{i,j} n_{ij} + \Sigma_{i,j} n_{ij} \bar{y}_{ij} \frac{(a_i + b_j - \frac{1}{\bar{y}_{ij}})^2}{2s^2})),$$

where $n = \Sigma_{i,j} n_{ij}$, $\bar{y}_{ij} = \Sigma_k \frac{y_{ijk}}{n_{ij}}$ and $n_{ij} = \Sigma_k (\frac{1}{y_{ijk}} - \frac{1}{\bar{y}_{ij}})$. The natural conjugate priors for α_i and β_j given σ^2 are found to be truncated normal distributions (Chhikara and Folks, 1989). The set of conjugate priors for all parameters are then given by

$$\alpha_i/\sigma^2 \sim \text{Normal}(0, \sigma^2 c^{-1}); \ \beta_j|\sigma^2 \sim \text{Normal}(\mu_b, \sigma^2 d^{-1}) I_{(b_j > a_i)};$$
$$\sigma^2 \sim \text{Inverse Gamma}\left(\frac{n_0}{2}, \frac{d_0}{2}\right); \ \mu_b|\sigma^2 \sim \text{Normal}(0, \sigma^2 f^{-1}) I_{(m_b > 0)};$$
$$c \sim \text{Gamma}\left(\frac{n_c}{2}, \frac{d_c}{2}\right); \ d \sim \text{Gamma}\left(\frac{n_d}{2}, \frac{d_d}{2}\right);$$
$$f \sim \text{Gamma}\left(\frac{n_f}{2}, \frac{d_f}{2}\right), \quad \text{.....(6)}$$

where $I_{(m_b > 0)}$ denotes the indicator function. Then by standard hierarchical Bayes calculations (see Gelfand and Smith, 1990), the full conditional distributions obtain as:

$[a_i|.]$=Normal $(\tilde{a}_i, \tilde{a}_i^{-1})$; $[b_j|.]$=Normal $(\tilde{b}_j, \tilde{b}_j^{-1})\, I_{(b_j>a_i)}$;

$[\sigma^2|.]$ = Inverse Gamma $(\frac{\tilde{n}_0}{2}, \frac{\tilde{d}_0}{2})$; $[m_b|.]$ = Normal $(\tilde{m}, \tilde{m}^{-1})\, I_{(m_b>0)}$;

$[c|.]$ = Gamma $(\frac{n_c+I}{2}, \frac{d_c+\sum_i \frac{a_i^2}{s^2}}{2})$;

$[d|.]$ = Gamma $(\frac{n_d+J}{2}, \frac{d_d+\sum_j \frac{(b_j-m_b)^2}{s^2}}{2})$;

$[f|.]$ = Gamma $(\frac{n_f+1}{2}, \frac{d_f+\frac{m_b^2}{s^2}}{2})$;

where,

$$\tilde{a}_i^{-1} = \frac{s^2}{(c+\sum_j n_{ij}\bar{y}_{ij})}, \quad \tilde{a}_i = \frac{\tilde{a}_i^{-1}\sum_j n_{ij}\bar{y}_{ij}(\bar{y}_{ij}^{-1}-b_j)}{s^2},$$

$$\tilde{b}_j^{-1} = \frac{s^2}{(d+\sum_j n_{ij}\bar{y}_{ij})}, \quad \tilde{b}_j = \tilde{b}_j^{-1}\left[\frac{dm_b}{s^2} + \frac{\sum_i n_{ij}\bar{y}_{ij}(\bar{y}_{ij}^{-1}-a_i)}{s^2}\right],$$

$$\tilde{m}^{-1} = \frac{1}{(f/s^2+dJ/s^2)}, \quad \tilde{m} = \frac{\tilde{m}^{-1}dJ\bar{b}}{s^2},$$

$$\tilde{d}_0 = d_0 + n + \sum_{ij} n_{ij}\bar{y}_{ij}(a_i + b_j - \bar{y}_{ij}^{-1}),$$
$$+ c\sum_i a_i^2 + d\sum_j (b_j - m_b)^2 + fm_b^2$$

$$\tilde{n}_0 = n_0 + I + J + 1, \quad n = \sum_{ij} n_{ij}, n_{ij} = \sum_k (\frac{1}{y_{ijk}} - \frac{1}{\bar{y}_{ij}}). \quad \text{.....(7)}$$

Each of the full conditional distributions has a simple form and is easily sampled from, thus providing via the Gibbs sampler, a complete sample from the joint posterior of the parameters. To generate from the constrained full conditional, we use Devroye's method (1986) or alternatively, sample from the unconstrained full conditional and retain the variate value only if it falls in the constraint region. Diffuse priors over (c, d, f, s^2) can be specified by setting the prior hyperparameters at their null-values, *i.e.*, $\nu_0 \to (-I-J-1)$, $\nu_c \to (-I)$ $\nu_d \to (-J)$, $\nu_f \to (-1)$, $\delta_0=0$, $\delta_c=0$, $\delta_d=0$, and $\delta_f=0$.

Posterior inference

The cell mean is $m_{ij} = \theta_{ij} = (\alpha_i+\beta_j)^{-1}$. Due to the nonlinearity of the cell mean parameters μ_{ij}, its posterior mean and variance are estimated (using (3)) via the sample averages as

$$\hat{E}(m_{ij}|\boldsymbol{y}) = N^{-1}\sum_{t=1}^{N}(a_i^{(t)} + b_j^{(t)})^{-1},$$

$$\hat{V}(m_{ij}|\boldsymbol{y}) = N^{-1}\sum_{t=1}^{N}(m_{ij}^{(t)} - \bar{m}_{ij})^2, \quad \text{.....(8)}$$

where $m_{ij}^{(t)} = (a_i^{(t)} + b_j^{(t)})^{-1}$ and $\bar{m}_{ij} = \hat{E}(m_{ij})$. The posterior densities of model parameters α_i, β_j, σ^2 and prior parameters μ_b, c, d, f are obtained using (2).

Predictive inference

Based on the posterior sample it is straightforward to derive the posterior predictive density of a future observation y_{ijk} and the mean $\bar{y}_{ij(m)}$ of m future observations from cell (i, j). Since $[y_{ijk}|y] = \int [y_{ijk}|\boldsymbol{y}, ?]\, d[?|\boldsymbol{y}]$, the posterior predictive density of y_{ijk} and $\bar{y}_{ij(m)}$ are estimated respectively, by finite mixtures as

$$[\hat{y}_{ijk}|\boldsymbol{y}] = N^{-1}\sum_{t=1}^{N} f(y_{ijk}|q_{ij}^{(t)}, s^{2(t)}),$$

$$[\hat{\bar{y}}_{ij(m)}|\boldsymbol{y}] = N^{-1}\sum_{t=1}^{N} f(\bar{y}_{ij(m)}|q_{ij}^{(t)}, s^{2(t)}), \quad \text{.....(9)}$$

where $f(y_{ijk})$ is the density of inverse Gaussian (θ_{ij}, s^2) and $f(\bar{y}_{ij(m)})$ is the density of inverse Gaussian $(\theta_{ij}, s^2/m)$ (see Fries and Bhattacharya,

1983). Furthermore, predictive means and variances of y_{ijk} and $\overline{y}_{ij(m)}$ can be easily calculated from the generated posterior sample. For instance, the predictive mean of y_{ijk} is $E[y_{ijk}|y] = \int \mathrm{E}[y_{ijk}|y,?]d[?|y]$, which is estimated as

$$\hat{E}[y_{ijk}|y] = N^{-1}\sum_{t=1}^{N} E(y_{ijk}|q_{ij}^{(t)}, s^{2(t)}) = N^{-1}\sum_{t=1}^{N} q_{ij}^{(t)}, \quad \text{.....(10)}$$

and the predictive variance is $V[y_{ijk}|y] = E(V(y_{ijk}|y,?)) + V(E(y_{ijk}|y,?))$, whose terms are estimated as

$$\hat{E}(V(y_{ijk}|y,?)) = N^{-1}\sum_{t=1}^{N} V(y_{ijk}|q_{ij}^{(t)}, s^{2(t)}) = N^{-1}\sum_{t=1}^{N} s^{2(t)} q_{ij}^{3(t)},$$

$$\hat{V}(E(y_{ijk}|y,?)) = (N-1)^{-1}\sum_{t=1}^{N}(d^{(t)} - \bar{d})^2, \quad \text{....(11)}$$

where $E(y_{ijk}|q_{ij}^{(t)}, s^{2(t)}) = q_{ij}^{(t)} = d^{(t)}$ say, and $\bar{d}$ is the average of the $d^{(t)}$. Similar calculations give the posterior predictive density of the mean $\overline{y}_{ij(m)}$ and its predictive mean and variance.

The lognormal model

We compare the inverse Gaussian model with the usual lognormal model for positively skewed data. The lognormal model is stated as

$y_{ijk} \sim$ lognormal (q_{ij}, s^2), $k = 1, \ldots, n_{ij}$, $\sigma^2 > 0$,
$E(y_{ijk}) = exp(\theta_{ij} + \sigma^2/2)$,
$V(y_{ijk}) = exp(2q_{ij} + \sigma^2)(exp(s^2) - 1)$. .....(12)

Let $z_{ijk} = \ln(y_{ijk})$. Then $z_{ijk} \sim$ normal (θ_{ij}, σ^2), $k = 1,\ldots,n_{ij}$. An additive model for factor effects is given by $\theta_{ij} = \alpha_i + \beta_j$, and the likelihood function for $(\{\alpha_i\}, \{\beta_j\}\ s^2)$ is

$$(2\pi\sigma^2)^{-n/2}\, exp\left(\Sigma_{i,j}(n_{ij} - 1)\frac{s_{ij}^2}{2s^2} + \Sigma_{i,j}\, n_{ij}\left(\frac{(\bar{z}_{ij} - a_i - b_j)^2}{2s^2}\right)\right),$$

where $n = \Sigma_{ij} n_{ij}$ $\bar{z}_{ij} = \Sigma_k z_{ijk}/n_{ij}$, and $s_{ij}^2 = \Sigma_k(z_{ijk}^{-1} - \bar{z}_{ij}^2)/(n_{ij} - 1)$. The natural conjugate priors for a_i and β_j are given by

$\alpha_i \sim$ Normal $(0, s_a^2)$; $\beta_j \sim$ Normal (m_b, s_b^2);
$\mu_b \sim$ Normal (m_0, s_0^2);
$s_a^2 \sim$ Inverse Gamma$(n_a/2, d_a/2)$;
$s_b^2 \sim$ Inverse Gamma$(n_b/2, d_b/2)$;
$s^2 \sim$ Inverse Gamma$(n_0/2, d_0/2)$. .....(13)

By standard hierarchical Bayes calculations, the full conditional distributions obtain as

$[a_i|.] =$ Normal $(\tilde{a}_i, \tilde{a}_i^{-1})$;
$[b_j|.] =$ Normal $(\tilde{b}_j, \tilde{b}_j^{-1})$;
$[m_b|.] =$ Normal $(\tilde{m}_0, \tilde{s}_0^2)$;
$[s_a^2|.] =$ Inverse Gamma $(\frac{n_a + I}{2}, \frac{(d_a + \Sigma_i a_i^2)}{2})$;
$[s_b^2|.] =$ Inverse Gamma $(\frac{n_b + J}{2}, \frac{d_b + \Sigma_j (b_j - m_b)^2}{2})$;
$[s^2|.] =$ Inverse Gamma $(\frac{n_0 + n}{2}, \frac{\tilde{d}_0}{2})$;

where,
$\tilde{a}_i^{-1} = [s_a^{-2} + n_i s^{-2}]^{-1}$, $\tilde{a}_i = \tilde{a}_i^{-1}[n_i(\bar{z}_i - \bar{z})s^{-2}]$,
$\tilde{b}_j^{-1} = [s_b^{-2} + n_j s^{-2}]^{-1}$, $\tilde{b}_j = \tilde{b}_j^{-1}[m_b s_b^{-2} + n_j \bar{z}_j s^{-2}]$,
$\tilde{s}_0^2 = [s_0^{-2} + J s_b^2]^{-1}$, $\tilde{m}_0 = \tilde{s}_0^2[m_0 s_0^{-2} + J\bar{b} s_b^{-2}]$,
$\tilde{d}_0 = d_0 + \sum_{ij}(n_{ij} - 1)s_{ij}^2 + \sum_{ij} n_{ij}(\bar{z}_{ij} - a_i b_j)^2$,
and $n = \sum_{ij} n_{ij}$. .....(14)

Diffuse priors over (m_b, s_a^2, s_b^2, s^2) can be specified by setting the prior hyperparameters at their null values as $m_0 = 0$, $n_a \to (-I)$, $n_b \to (-J)$, $n_0 \to (-n)$, $s_0^2 = 10^{12}$, $d_a = 0$, $d_b = 0$ and $d_0 = 0$. The cell mean parameter is $\mu_{ij} = exp(\theta_{ij} + \sigma^2/2)$. Using (4), its posterior mean is estimated as

$$\hat{E}(m_{ij}|y) = N^{-1}\sum_{t=1}^{N} \exp(a_i^{(t)} + \tfrac{s^{2(t)}}{2}) \exp(\tilde{b}_j^{(t)} + \tfrac{\tilde{b}_j^{-1(t)}}{2}), \quad \text{.....(15)}$$

where $\tilde{b}_j^{(t)}$, and $\tilde{b}_j^{-1(t)}$ are the mean and variance of β_j obtained from its full conditional distribution. Its posterior variance $V(\mu_{ij}|y) = E(V(\mu_{ij}|y, \eta) + V(E(\mu_{ij}|y, \eta))$, η being the hyperparameters, is estimated by the sum of the two components

$$\hat{E}(V(m_{ij}|\mathbf{y},h)) = N^{-1}\sum_{t=1}^{N}\exp(2a_i^{(t)} + s^{2(t)})\exp(2\tilde{b}_j^{(t)} + \tilde{b}_j^{-1(t)})(\exp(\tilde{b}_j^{-1(t)}) - 1),$$

$$\hat{V}(E(m_{ij}|\mathbf{y},h)) = (N-1)^{-1}\sum_{t=1}^{N}(d^{(t)} - \bar{d})^2, \quad \text{.....(16)}$$

where

$$d^{(t)} = N^{-1}\sum_{t=1}^{N}\exp(a_i^{(t)} + \frac{s^{2(t)}}{2})\exp(\tilde{b}_j^{(t)} + \frac{\tilde{b}_j^{-1(t)}}{2}),$$

and $\bar{d}$ is the average of the $d^{(t)}$. The posterior densities of model parameters α_i, β_j, σ^2 and prior parameters m_b, s_a^2, s_b^2 are obtained using (2).

The posterior predictive density of a future observation y_{ijk} is $[y_{ijk}|y] = \int[y_{ijk}|y, q]d[q|y]$ which is estimated as a finite mixture

$$[\hat{y}_{ijk}|y] = N^{-1}\sum_{t=1}^{N} f(y_{ijk}|q_{ij}^{(t)}, s^{2(t)}), \quad \text{.....(17)}$$

where $f(y_{ijk})$ is the density of lognormal (θ_{ij}, s^2). Furthermore, the predictive mean and variance of y_{ijk} are estimated using formulas similar to (10) and (11). The predictive density of the mean $\bar{y}_{ij(m)}$ of m future observations and its predictive mean and variance are estimated in a similar fashion, except $f(\bar{y}_{ij(m)})$ is defined as the large-m normal approximation with mean $exp(q_{ij}+s^2/2)$ and variance $m^{-1}\, exp(q_{ij}+s^2/2)(exp(s^2)-1)$.

Estimates for finite populations

Consider a finite population U_{ij} with units labelled $1, \ldots, N_{ij}$. Let y_{ijk} denote the value of a single characteristic attached to the unit k in population U_{ij}. The vector $\{y_{ijk}; k = 1, \ldots, N_{ij}\}$ is the unknown state of nature. We assume that the population of N_{ij} elements for area (i, j) is generated by the super population model $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}; k = 1, \ldots, N_{ij}$. Assume also that n_{ij} observations are available for area (i, j). The mean of the n_{ij} observations is the observed mean $\bar{y}_{ij(n_{ij})}$ and the mean of the unobserved $(N_{ij} - n_{ij})$ elements, $\bar{y}_{ij(N_{ij}-n_{ij})}$, and its variance are estimated by the Bayes predictors similar to the results (9) through (11) for inverse Gaussian and lognormal errors. Letting $f_{ij} = n_{ij}/N_{ij}$, we obtain the estimator of the finite population mean $\bar{y}_{ij(N_{ij})}$ as

$$\hat{\bar{y}}_{ij(N_{ij})} = f_{ij}\bar{y}_{ij(n_{ij})} + (1 - f_{ij})\hat{\bar{y}}_{ij(N_{ij}-n_{ij})}. \quad \text{.....(18)}$$

APPLICATIONS

The method is applied to the 1986 Canadian data on household income obtained from Statistics Canada (1987). Subsets of the population were obtained by geographic region and education status of the head of household. The variable of interest, total income (y_{ijk}), is defined as the sum of total earnings, total income from investment, total government transfer payments, retirement pensions, superannuation and annuities and other money income. This variable may be positive, negative or zero. For the purpose of conducting the analysis, only cases with positive values were retained and all others were discarded. Out of a sample of 30,741 there were exactly 98 cases with non-positive values. The reduced sample size of positive sample values was 30,643.

The geographic regions were chosen as the ten provinces of Canada: (1) Newfoundland, (2) Prince Edward Island, (3) Nova Scotia, (4) New Brunswick, (5) Quebec, (6) Ontario, (7) Manitoba, (8) Saskatchewan, (9) Alberta, and (10) British Columbia. The education classes were defined as (1) No schooling or elementary, (2) 9 or 10 years of elementary and secondary, (3) 11–13 years of elementary and secondary, (4) Some post-secondary, (5) Post-secondary certificate or diploma, and (6) University degree. The parameters α_i, $i = 1, \ldots, 6$; β_j, $j = 1, \ldots, 10$, represent the effects due to the six education levels and the 10 provinces, respectively.

Table A in the Appendix shows the cross-classification of the sample into six education classes and 10 provinces with corresponding cell counts and cell means. Note that sample sizes are of order not exceeding 1000. For the purpose of illustrating the two models in finite population sampling theory, a 10% random sample was selected from each of the 10×6 = 60 subpopulations.

Inverse Gaussian errors

Under the inverse Gaussian model, we rescale the data by multiplying by a factor of 10^{-5} in the Bayesian computations to maintain numerical stability. The prior cell mean m_b^{-1} is estimated as $N^{-1}\sum_{t=1}^{N} m_b^{-1(t)}$. The mean-variance structure of the model is assessed for each cell (i, j) via the goodness-of-fit statistic

$$\hat{c}_{ij}^{\,2} = \frac{(\overline{y}_{ij} - \hat{E}(\overline{y}_{ij}))^2}{\hat{V}(\overline{y}_{ij})} \sim c^{\,2}(1), \quad \text{.....(19)}$$

where

$$\hat{E}(\overline{y}_{ij}) = N^{-1}\sum_{t=1}^{N} q_{ij}^{(t)},$$

$$\hat{V}(\overline{y}_{ij}) = n_{ij}^{-1}N^{-1}\sum_{t=1}^{N} s^{\,2(t)} q_{ij}^{3(t)} \quad \text{.....(20)}$$

We used Matlab in the development of the algorithm and implementation of the Gibbs sampler. Six hundred Gibbs sequences with different starting values were sampled until the 1,000-*th* iteration in multiple runs of the Gibbs sampler due to the nonlinearity of the parameters and due to the constraint that $\alpha_i + \beta_j > 0$. These restrictions gave a poor parametrization of the Gibbs sampler which in turn caused poor convergence of a single run (see Hills and Smith, 1992). The hyperparameters in (7) are set at their null values to give vague priors. Figures 1 and 2 display the kernel density estimates of the sampled posteriors for α_i, i = 1(1)6, and for β_j, j = 1(1)10, under the inverse Gaussian model. Figure 1 clearly shows the decreasing rank order of the α_i which on the reciprocal scale gives an increasing rank order of $(1/\alpha_i)$ (education effect). This agrees with the typical situation that as education increases, income also increases. As for the β_j, j = 1(1)10, parameters from Figure 2, we can group them by their location parameters to detect provinces with similar income levels.

The posterior means and standard deviations, obtained using (8), along with the goodness-of-fit statistic for each cell based on (19) and (20) are presented in Table 1 for the inverse Gaussian model. Comparisons of these cell means with the observed cell means of Table A of the Appendix reveal that the amount of shrinkage of observed means toward the prior mean is considerable for the smaller sized subpopulations.

Table 1. Small area estimation under the inverse Gaussian model: estimated posterior cell means, standard deviations and goodness-of-fit statistics.

EDUCATION	PROVINCE									
	Posterior ell means: $\hat{E}(\overline{y}_{ij})$									
	1	2	3	4	5	6	7	8	9	10
1	22039	22224	22823	23251	24253	26183	23376	23643	25320	24333
2	26226	26488	27343	27959	29420	32308	28138	28526	31003	29536
3	29163	29490	30549	31321	33165	36881	31548	32034	35191	33312
4	30092	30440	31572	32395	34372	38380	32636	33160	36555	34530
5	31855	32246	33521	34450	36691	41293	34724	35314	39190	36873
6	41271	41931	44099	45720	49751	58621	46208	47263	54458	50098
	Posterior standard deviations: $\hat{V}(\overline{y}_{ij})$									
	1	2	3	4	5	6	7	8	9	10
1	397	471	349	392	290	367	369	333	369	363
2	591	677	546	598	488	575	533	489	529	544
3	694	839	610	665	512	607	651	540	556	569
4	861	988	849	866	793	934	804	812	891	830
5	884	1051	898	938	783	897	906	806	906	861
6	1510	1755	1411	1474	1243	1761	1497	1464	1483	1560
	Goodness-of-fit statistics: $\hat{c}_{ij}^{\,2}$									
	1	2	3	4	5	6	7	8	9	10
1	3.215	1.126	3.355	.067	2.270	6.001	.032	0.481	.037	.182
2	.426	.615	.368	.117	.083	2.881	.072	.251	0.264	1.075
3	.037	.085	.458	.477	2.458	1.147	.069	.427	.011	2.345
4	.292	.422	2.219	.068	.072	.229	.331	.736	1.331	.697
5	2.242	1.028	.037	0.438	006	1.600	0.002	.680	.004	.001
6	1.609	1.545	.468	.116	.004	1.718	0.034	2.826	0.039	0.879

Lognormal errors

For the lognormal errors, the model was reparameterized to the logarithmic scale for efficient derivation of the Gibbs sampler. All posterior densities are displayed on the logarithmic scale. The prior cell mean $exp(\mu_b+\sigma^2)$ is estimated as $N^{-1}\sum_{t=1}^{N}\exp(m_b^{(t)}+s^{2(t)})$.

A goodness-of-fit statistic to assess the mean-variance structure of the model for each cell (i, j) is given by (19), where now

$$\hat{E}(\overline{y}_{ij}) = N^{-1}\sum_{t=1}^{N}\exp(q_{ij}^{(t)}+\tfrac{s^{2(t)}}{2}),$$

$$\hat{V}(\overline{y}_{ij}) = n_{ij}^{-1}N^{-1}\sum_{t=1}^{N}\exp(2q_{ij}^{(t)}+s^{2(t)})(\exp(s^{2(t)})-1). \quad \cdots(21)$$

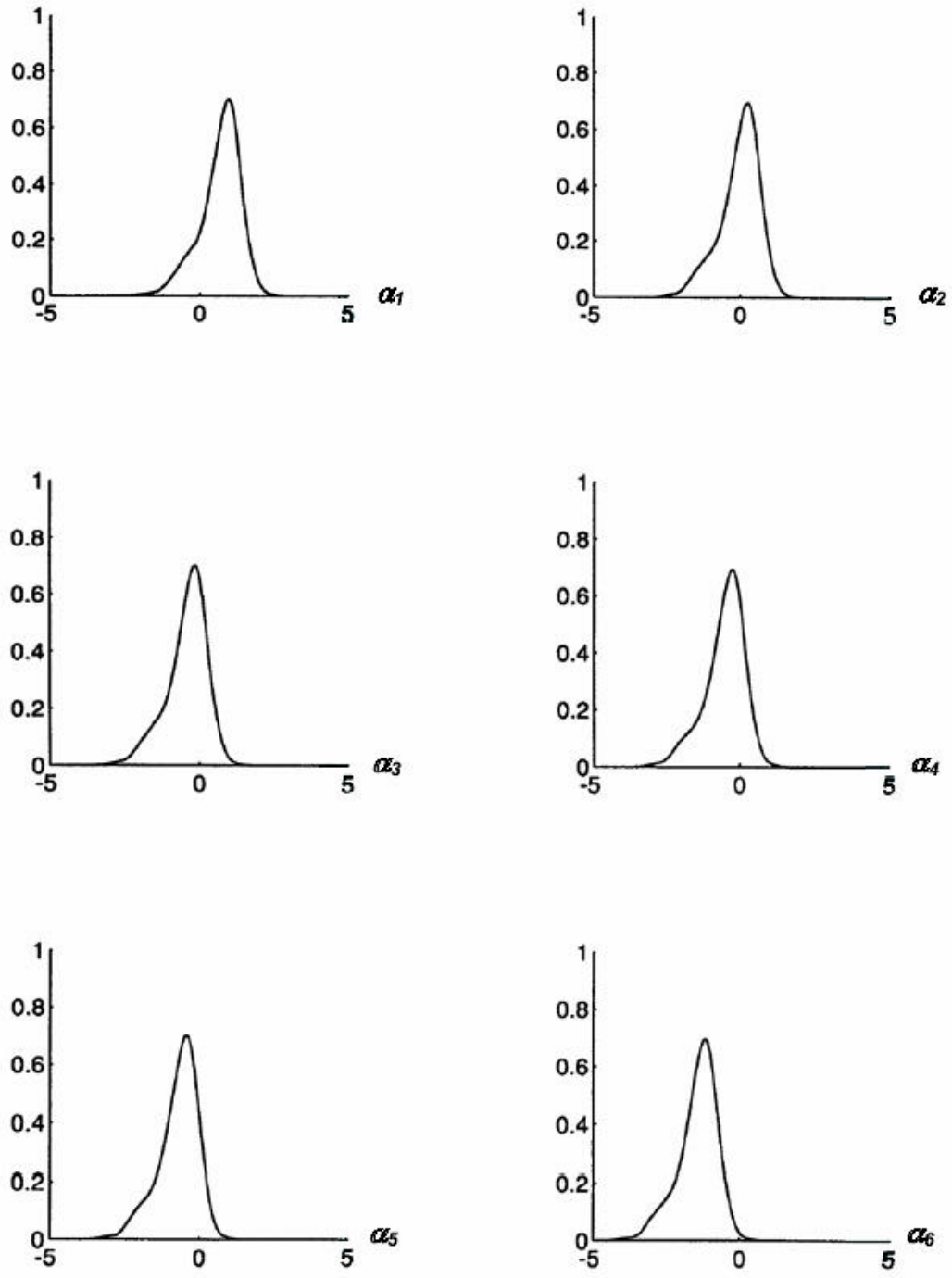

Fig. 1. Kernel density estimates of the sampled posteriors for a_i, $i=1(1)6$, under the inverse Gaussian model.

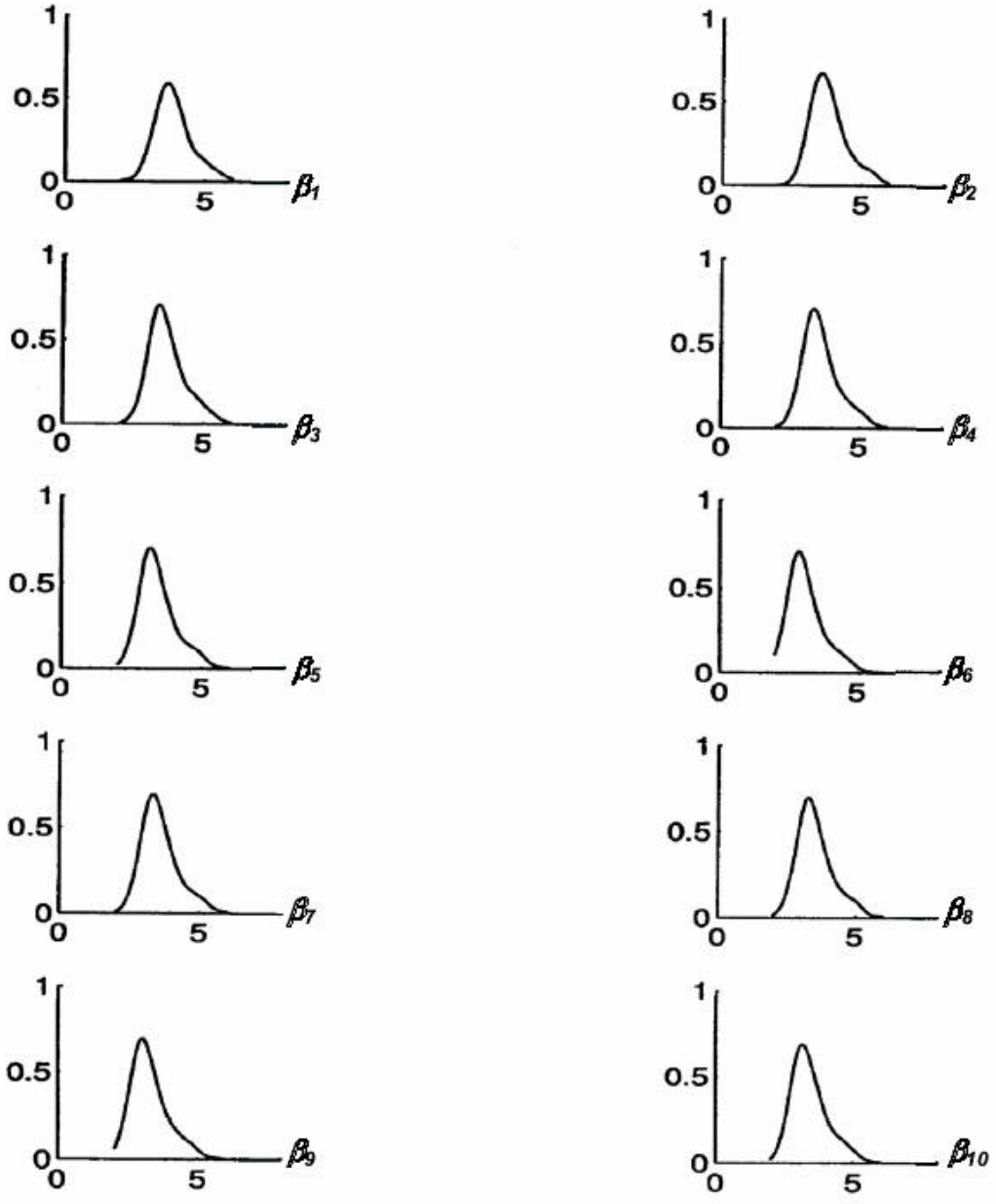

Fig. 2. Kernel density estimates of the sampled posteriors for b_j, j=1(1)10 , under the inverse Gaussian model.

In the implementation of the Gibbs sampler, the first 2,000 draws of a single run are discarded and the algorithm is run to obtain 1,000 draws from the posterior. The results for this case are presented in Table 2, where the estimated posterior cell means, posterior standard deviations and the goodness-of-fit statistics are given. The hyperparameters in (14) are set at their null values to give vague priors. Displays of the kernel density estimates of the sampled posteriors for α_i, i = 1(1)6, and for β_j, j = 1(1)10, under the lognormal model are not included in the paper in the interest of saving space. The results are similar to those of the inverse Gaussian case but with less pronounced effects (Chaubey *et al.*, 2003).

Table 2 displays estimated posterior means and standard deviations obtained using (8) along with the goodness-of-fit statistic for each cell based on (19 and 21) for the lognormal model. The results are parallel to those given in Table 1 for the inverse Gaussian model.

Table 2. Small area estimation under the lognormal model: estimated posterior cell means, standard deviations and goodness-of-fit statistics.

EDUCATION	PROVINCE									
	Posterior ell means: $\hat{E}(\overline{y}_{ij})$									
	1	2	3	4	5	6	7	8	9	10
1	20832	21307	21880	22602	23545	27755	22284	22264	26292	24922
2	25779	26367	27076	27970	29137	34347	27576	27551	32535	30840
3	30325	31017	31850	32902	34275	40404	32439	32409	38273	36279
4	30701	31401	32245	33310	34700	40904	32841	32811	38747	36729
5	34331	35114	36057	37248	38802	45740	36723	36690	43328	41071
6	46832	47900	49187	50811	52931	62396	50096	50050	59105	56026
	Posterior standard deviations: $\hat{V}(\overline{y}_{ij})$									
	1	2	3	4	5	6	7	8	9	10
1	403	533	374	385	316	345	376	338	357	375
2	525	681	494	509	434	480	498	454	488	505
3	578	770	535	551	447	485	538	481	505	534
4	722	891	700	722	658	747	707	664	739	740
5	738	939	704	725	636	713	710	656	716	729
6	1007	1280	960	988	867	972	967	894	976	994
	Goodness-of-fit statistics: $\hat{c}_{ij}^{2}$									
	1	2	3	4	5	6	7	8	9	10
1	.001	.013	.193	1.219	7.479	.006	2.353	.111	1.168	.048
2	.061	.471	.128	.137	0.001	0.236	.056	2.337	1.381	6.480
3	.359	1.422	.337	.573	12.728	59.869	1.588	.104	23.048	4.357
4	.126	.144	1.947	.601	.027	2.581	.354	.706	11.753	7.699
5	.067	.054	1.369	6.288	2.779	44.307	1.554	.006	17.059	10.149
6	.023	.001	2.251	9.376	6.226	44.458	2.622	1.695	10.669	32.869

DISCUSSION

The results from Table 1 under the inverse Gaussian model, and from Table 2 for the lognormal model, show that the 'education effect' reflects the typical situation that as education increases, income also increases. Comparisons of the posterior cell means from these tables with the actual cell means from Table A in the Appendix show varying amounts of shrinkage of observed means towards the prior means and that the shrinkage is more for the smaller sized subpopulations (higher education levels and smaller provinces). The shrinkages are more pronounced for the inverse Gaussian model than for the lognormal model. Further more, the inverse Gaussian model appears to give a better fit to the data than the lognormal model. From Table 1 under the inverse Gaussian case, that at a 5% significance level for which $c_{.05}^{2}(1) = 3.84$, the fit is assessed to be poor only for a single cell ($i = 1, j = 6$), whereas in Table 2 under the lognormal case, the fit is assessed to be poor for almost 25% of the cells. These results are confirmed from Figure 3, where the Chi-squares goodness-of-fit statistics under the two models are compared.

Figure 4 displays the posterior predicted finite population cell means under the two models in relation to the observed cell means from Table A, based on a 10% sample. We observe again that the inverse Gaussian model provides estimates closer to the true values than the lognormal model for this application. Furthermore, we note that education classes with relatively small sample sizes appear to give less reliable statistics in both cases.

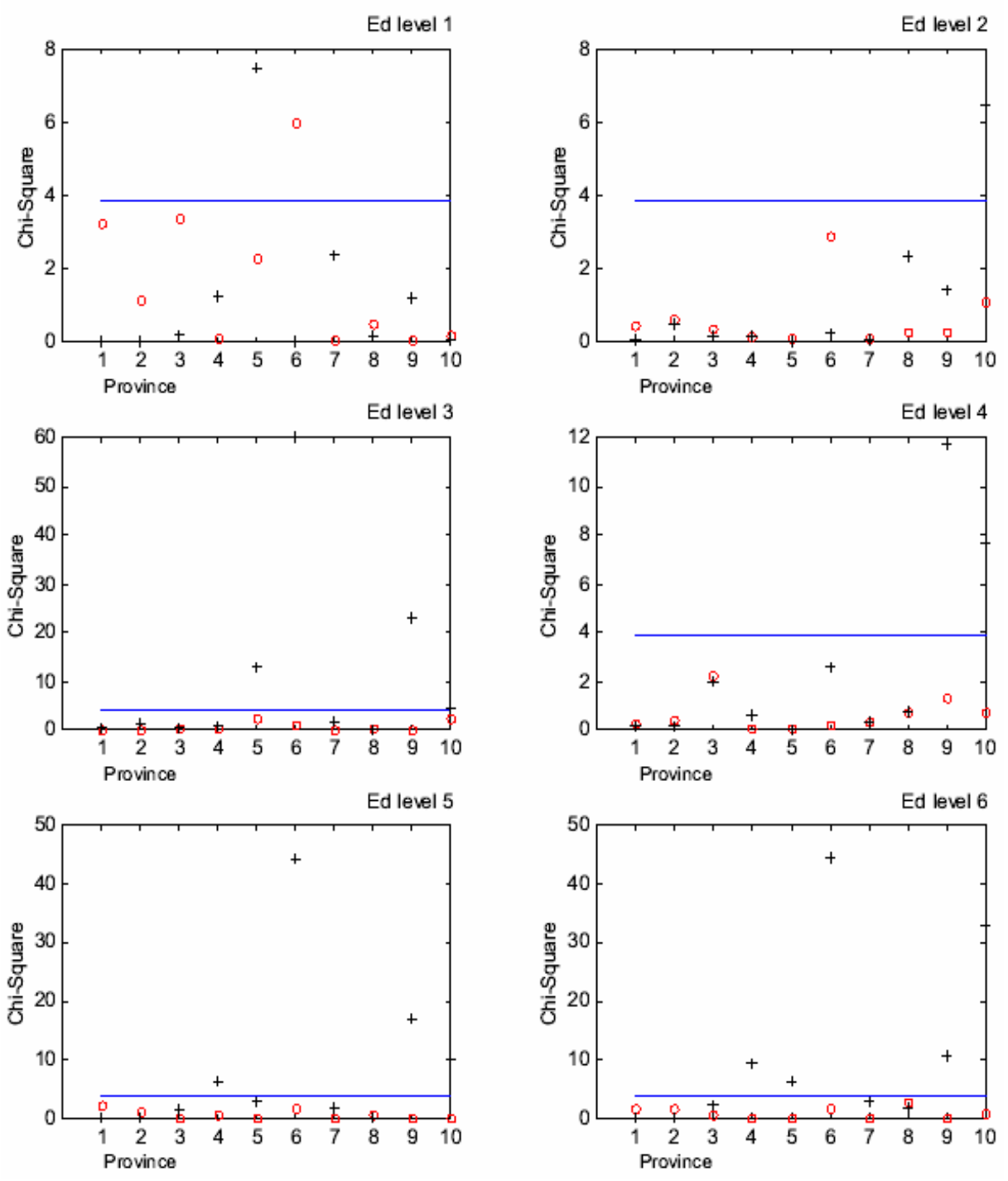

Fig. 3. Chi-squares goodness-of-fit statistics from Tables 1 and 2 for inverse Gaussian and lognormal models for Provinces by Levels of Education (Ed Level i, i =1(1)6): Inverse Gaussian (o); Lognormal (+); $\hat{c}^{\,2}_{1,0.05}$ = 3.84 (–).

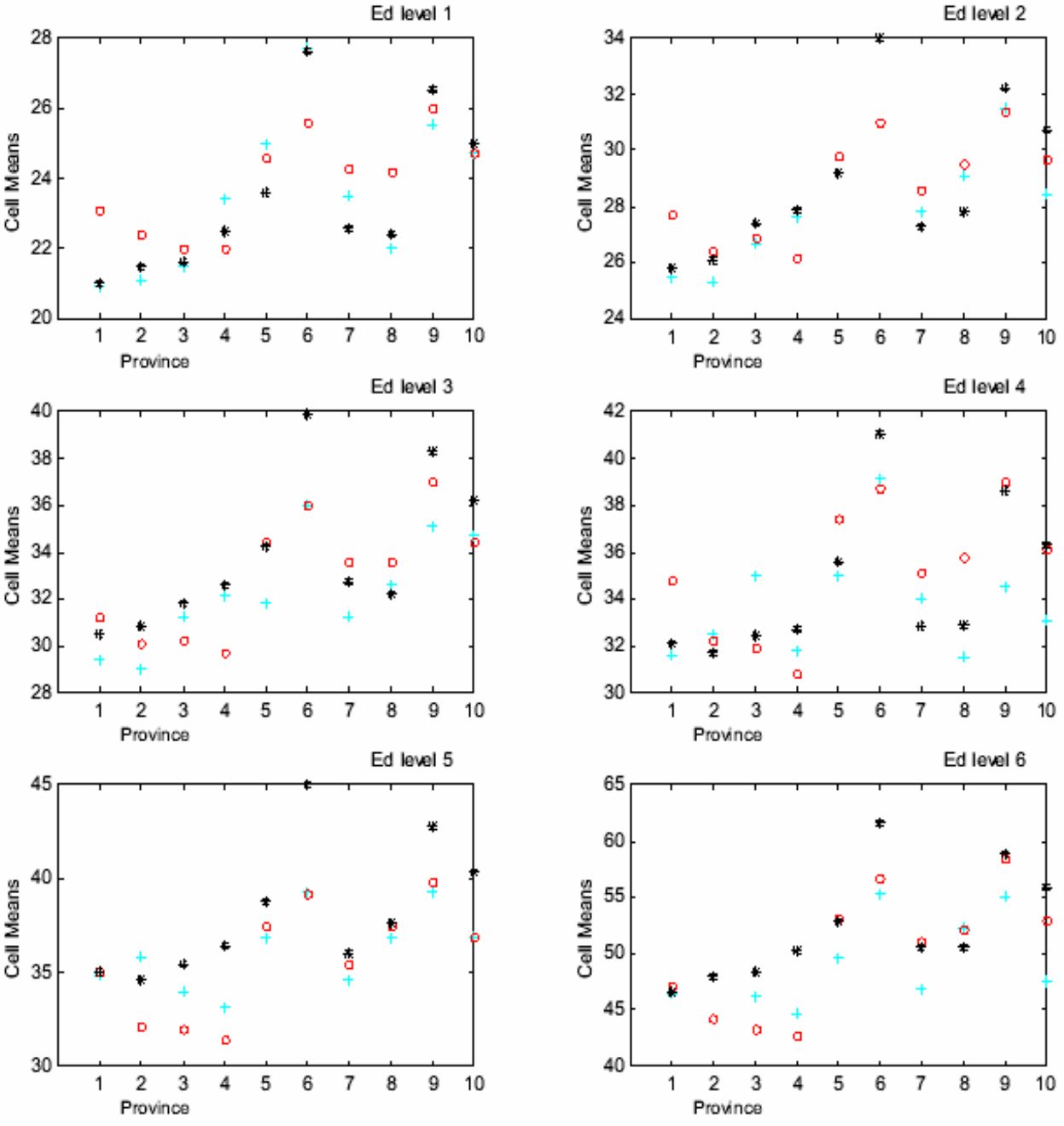

Fig. 4. Posterior predicted finite population means for *Provinces by Levels of Education* (*Ed Level i*, i =1(1)6): Observed Means $\bar{y}_{ij}$ (o); Estimated Means $\hat{m}_{ij}$ under Inverse Gaussian (+) and Lognormal (*) models.

CONCLUSION

We have used Gibbs sampling to obtain estimates of small area parameters under two kinds of non-normal sampling errors. One can see that it is possible to obtain estimates of small area parameters using the Bayesian approach by borrowing strength from an ensemble when the sampling models are non-normal. It is evident from the illustrations provided based on the household survey data that the inverse Gaussian model appears to be more suitable for modeling positively skewed data compared to the lognormal model. Thus, the possible use of inverse Gaussian model should be explored side by side with the lognormal model.

ACKNOWLEDGEMENTS

Funding from NSERC to Y.P. Chaubey and seed grants from NSERC and Health Canada to Fassil Nebebe are gratefully acknowledged. The analysis is based on household income from the 1987 Statistics Canada microdata tape. All computations on these microdata were done by the authors and the responsibility for the use and interpretations of these data are entirely that of the authors. The authors are also thankful to the anonymous reviewers for their helpful comments.

REFERENCES

1. Bhattacharyya, G.K. and Fries, A. (1982). Inverse Gaussian regression and accelerated life tests. In: *Survival Analysis,* pp. 101–117, (Crowley, J. and Johnson, R.A., eds). Institute of Mathematical Statistics, Hayward, CA.
2. Casella, G. and George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician* **46**:167–174.
3. Chaubey, Y.P., Fassil Nebebe and Chen, P.S. (1994). Small area estimation under inverse Gaussian model. *Survey Methodology* **22**:33–41.
4. Chaubey, Y.P., DeSouza, C.M. and Fassil Nebebe (2003). Bayesian inference for small area estimation under the inverse Gaussian model via Gibbs sampling. *Technical Report No. 4/03*, Department of Mathematics and Statistics, Concordia University, Montreal, Quebec, H4B1R6 Canada.
5. Chhikara, R.S. and Folks, J.L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications.* Marcel Dekker, New York.
6. Cowles, M.K. and Carlin, B.P. (1996). Markov Chain Monte Carlo convergence diagnostics: A Comparative Review. *Journal of the American Statistical Association* **91**:883–904.
7. Cox, D.R. and Miller, H.D. (1965). *The Theory of Stochastic Processes.* Chapman and Hall, New York, pp. 210–212.
8. Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer Verlag, New York, pp. 27–77.
9. Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* **74**:269–277.
10. Fries, A. and Bhattacharyya, G.K. (1983). Analysis of two-factor experiments under an inverse Gaussian model. *Journal of the American Statistical Association* **78**:820–826.
11. Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**:972–985.
12. Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**:398–409.
13. Gelfand, A.E. and Smith, A.F.M. (1991). Gibbs sampling for marginal posterior expectations. *Communications in Statistics* **B(20):**1747–1766.
14. Gelfand, A.E., Smith, A.F.M. and Lee, T.M. (1992). Bayesian analysis of constrained parameters and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association* **87**:523–532.
15. Gelman, A. and Rubin, D.B. (1992). Inference from interactive simulation using multiple sequences. *Statistical Science* **7**:457–472.
16. Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimators in finite population sampling. *Journal of the American Statistical Association* **81**:1058–1062.
17. Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: an appraisal. *Journal of the American Statistical Association* **9**:55–93.
18. Hills, S.E. and Smith, A.F.M. (1992). Parametrization issues in Bayesian inference. *Bayesian Statistics* **4**:227–246.
19. Malec, D., Davis, W.W. and Cao, X. (1999). Model-based small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine* **18**:3189–3200.
20. MacGibbon, B. and Tomberlin, T.J. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology* **15**:237–252.
21. Purcell, N.J. and Kish, L. (1979). Estimation for small domains. *Biometrics* **35**:365–384.
22. Rao, J.N.K. (2003). *Small Area Estimation.* Wiley, New York.
23. Rao, J.N.K. (1999). Some recent advances in model based small area estimation. *Survey Methodology* **25**:175–186.
24. Roberts, G.O. and Rosenthal, J.S. (1998). Markov-Chain Monte Carlo: some practical implications of theoretical results. *Canadian Journal of Statistics* **26**:5–31.
25. Schaible, W.L. (1996). *Indirect Estimation in U.S. Federal Programs*. Springer, New York.
26. Statistics Canada (1987). *Microdata File Documentation on Household Income (1986), Facilities and Equipment (1987).* Statistics Canada, Household Survey Division, Ottawa, Canada.
27. Stroud, T.W.F. (1991). Hierarchical Bayes predictive means and variances with application to sample survey inference. *Comm. Statist. Theory and Methods* **20**:13–36.

Appendix : Table A. Actual cell counts and cell means.

EDUCATION [b]	PROVINCE [a]									
	Cell counts n_{ij}									
	1	2	3	4	5	6	7	8	9	10
1	626	285	597	729	1371	1177	636	841	698	456
2	360	212	483	396	760	883	430	507	695	540
3	277	197	616	567	1212	1793	672	877	1338	1081
4	84	72	148	150	202	516	164	263	382	341
5	215	68	203	239	471	704	233	349	693	389
6	110	83	230	219	508	800	222	297	570	406
	Cell means $\overline{y}_{ij}$									
	1	2	3	4	5	6	7	8	9	10
1	20856	21173	21519	23423	25028	27709	23505	22033	25470	24716
2	25488	25315	26713	27553	29152	33982	27831	29067	31540	28394
3	29453	28959	31284	32132	31792	35977	31261	32673	35094	34741
4	31646	32493	35041	31767	34981	39181	33973	31548	34472	33066
5	34786	35841	33941	33067	36822	39275	34619	36792	39284	36908
6	46394	47854	46209	44583	49599	55300	46825	52324	54987	47462

[a]PROVINCE: 1=Newfoundland; 2=Prince Edward Island; 3=Nova Scotia; 4=New Brunswick; 5=Quebec; 6=Ontario; 7=Manitoba; 8=Saskatchewan; 9=Alberta; 10=British Columbia.

[b]EDUCATION: 1=No schooling or elementary; 2=9 or 10 years of elementary and secondary; 3=11–13 years of elementary and secondary; 4=some post secondary; 5=post secondary certificate or diploma; 6=university degree.